7.0 CONCLUSIONS AND GUIDANCE TO LABORATORIES, PERMITTEES, AND REGULATORY AUTHORITIES

This document was prepared to address whole effluent toxicity (WET) test variability. The document has three goals: (1) quantify the variability of promulgated test methods and report a coefficient of variation (CV) as a measure of test method variability; (2) evaluate the statistical methods described in the TSD for determining the need for and deriving WET permit conditions; and (3) suggest guidance for regulatory authorities on approaches to address and minimize test method variability. This document quantified the variability of toxicity test methods based on the end use of the data, that is, the effect concentrations (e.g., NOEC, LC50, EC25). The within-laboratory variability of these effect concentrations was quantified by obtaining multiple test results under similar test conditions using the same reference toxicant. The major conclusions of this document are discussed below.

7.1 General Conclusions

- EPA's *Technical Support Document for Water Quality-based Toxics Control* (referred to as the TSD) presents guidance for developing effluent limits based on three key components: (1) water quality criteria; (2) a calculated dilution factor used to derive a waste load allocation (WLA) from the criteria; and (3) a statistical calculation procedure that uses a CV based on effluent data to calculate effluent limits from the WLA. EPA's TSD statistical approach is appropriately protective, regarding both effluent and analytical variability, provided that the criteria and WLA are derived correctly. It is inappropriate to adjust the TSD statistical methodology for determining when water quality-based effluent limits are needed and for calculating such limits (Section 6 and Appendix G).
- EPA's analysis indicates that the TSD approach appropriately accounts for both effluent variability and method variability. EPA does not believe a reasonable alternative approach is available to determine a factor that would discount the effects of method variability using the TSD procedures (Section 6.1.1 and Appendix G).
- Interim CVs are identified for promulgated WET test methods [Appendix A, Table A-1 (acute methods) and Table A-2 (chronic methods)], pending completion of between-laboratory studies, which may affect these interim CV estimates.
- Comparisons of WET method precision with method precision for analytes commonly limited in the National Pollutant Discharge Elimination System (NPDES) permits clearly demonstrate that the variability of the promulgated WET methods is within the range of variability experienced in other types of analyses. Several researchers also noted that method performance improves when prescribed methods are followed closely by experienced analysts (Section 4.3).
- The hypothesis test procedures prescribed in EPA's WET methods will provide adequate protection against false conclusions that an effluent is toxic. However, the incidence of false negatives can be high because of high within-test variability, making it difficult to detect toxicity when toxicity is truly present. Therefore, evaluating the power of current experimental designs is desirable. EPA expects that regulatory authorities will make prompt and measurable progress toward the goal of requiring all WET tests to detect a toxic effect of 25 percent to 33 percent with power of 0.80 (Section 5.3.3 and Appendix B.4).
- Quality assurance problems became apparent when evaluating the data for this study, especially for the metal reference toxicants and sodium dodecyl sulfate (SDS). Standardizing the choice of reference toxicant and the concentrations to be tested may be appropriate, as well as establishing bounds on the range of acceptable effect concentrations for each test method. As a result,

quantifying between-laboratory variability will be difficult unless these issues can be resolved (Section 5.3.1 and Appendix G.2.5).

- The data analysis did not reveal the potential sources and causes of variability, such as using different sources of test organisms, dilution water, and food. To assess the sources of variability fully, experimenters must carefully design new studies (Section 5.3.1).

7.2 Recommendations for Minimizing Variability and Its Effects

Three critical areas are identified to minimize WET test method variability:

- Obtaining a representative effluent sample,
- Conducting the toxicity tests properly to generate biological endpoints, and
- Calculating the appropriate statistical endpoints to have confidence in the effect concentration.

This document provides guidance to toxicity testing laboratories, permittees, and regulatory authorities in conducting biological and statistical methods and evaluating test effect concentrations. It also develops guidance for regulatory authorities on approaches to address and minimize test method variability. The principal aspects of the guidance are presented in Table 3-6 and re-presented here.

Range of Relative Variability for Endpoints of Promulgated WET Methods, Defined by the 10th and 90th Percentiles from the Data Set of Reference Toxicant Tests[a]

Test Method[b]	Endpoint[c]	No. of Labs	No. of Tests	PMSD		Control CV[d]	
				10th	90th	10th	90th
1000.0 Fathead Minnow	G	19	205	9.4	35	0.035	0.20
1002.0 *Ceriodaphnia dubia*	R	33	393	11	37	0.089	0.42
1003.0 Green Alga	G	9	85	9.3	23	0.034	0.17
1004.0 Sheepshead Minnow	G	5	57	6.3	23	0.034	0.13
1006.0 Inland Silverside	G	18	193	12	35	0.044	0.18
1007.0 Mysid	G	10	130	12	32	0.088	0.28
2000.0 Fathead Minnow	S	20	217	4.2	30	0	0.074
2002.0 *Ceriodaphnia*	S	23	241	5.0	21	0	0.11
2004.0 Sheepshead Minnow	S	5	65	0[e]	55	0	0
2006.0 Inland Silverside	S	5	48	7.0	41	0	0.079
2007.0 Mysid (*A. bahia*)	S	3	32	5.1	26	0	0.081
2011.0 Mysid (*H. costata*)	S	2	14	18	47	0	0.074
2021.0 Daphnia (*D. magna*)	S	5	48	5.3	23	0	0.11
2022.0 Daphnia (*D. pulex*)	S	6	57	5.8	23	0	0.11

[a] The precision of the data warrants only three significant figures. When determining agreement with these values, one may round off values to two significant figures (e.g., values >3.45000... and ≤3.5000... are rounded to 3.5). Method 1009.0 (red macroalga) is not reported because it is inadvisable to characterize method variability using only 23 tests from just two laboratories.

[b] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

[c] G = growth, R = reproduction, S = survival

[d] CVs were calculated using untransformed control means for each test.

[e] An MSD of zero will not occur when the EPA flow chart for statistical analysis is followed. In this report, MSD was calculated for every test, including those for which the flow chart would require a nonparametric hypothesis test. EPA recommends using the value 4.2 (the 10th percentile shown for the fathead minnow acute test) in place of zero as the 10th percentile PMSD (lower PMSD bound) for the sheepshead minnow acute test.

7.2.1 Guidance to Toxicity Testing Laboratories

- Testing laboratories should maintain quality assurance/quality control (QA/QC) control charts for percent minimum significant difference (PMSD) along with the statistical endpoints such as NOEC, LC50, and EC25. Testing laboratories should regularly plot the individual raw test data and the average treatment responses to examine possible causes of excessive variability (Section 5.3.1.1).

- The minimum number of replicates for the chronic toxicity tests should be four for the chronic fathead minnow, sheepshead minnow, and inland silverside test methods (Sections 5.3.3.1 and 5.6).

- Testing laboratories should take steps to ensure that the test PMSD does not exceed the upper bound provided in the table above (Sections 3.3, 5.3.3, and 6.4 and Table 3-6). This may require ensuring more uniformity among test organisms and/or using more replicates. Tables are provided to aid in choosing the number of replicates (Tables B-14 and B-15).

- Testing laboratories should examine the power tables to ensure that test results will meet the recommended test sensitivity criteria. These tables can be used to make decisions about replication, given the knowledge of typical values for error mean square (EMS) and number of tested concentrations (Section 5.3.3 and Tables B-9 through B-15).

7.2.2 Guidance to NPDES Permittees

- Permittees should select and conduct all data analyses with one qualified toxicity testing laboratory to determine reasonable potential, derive permit limits, and generate self-monitoring test results. Conducting all effluent testing consistently using one reference toxicant is also prudent (Section 6.1.4 and Appendix G.2.5).

- Permittees should generate WET data (n = 10) that have been accumulated over a year or more to fully characterize effluent variability over time. The sampling dates and times should span a sufficient duration to represent the full range of effluent variability (Sections 6.1.3 and 6.2 and Appendix G.2.4).

- Permittees should examine testing laboratories' QA/QC control charts. If the CV for reference toxicant tests is greater than the 75^{th} percentile in Tables 3-2 through 3-4, variability can likely be reduced, even if the individual EC25 and LC50 values fall within the control limits (Section 5.3.1.1).

- Permittees should examine toxicity test data to ensure that data being submitted to regulatory authorities meet specified effluent holding times, temperature, laboratory control limits, and test acceptability criteria, such as requirements for test sensitivity lower and upper PMSD bounds (Sections 5.2 through 5.4).

- Permittees should anticipate and plan for a change if switching to a different testing laboratory. The permittee should compare reference toxicant test data from the current laboratory with data from the candidate replacement laboratory in order to ensure acceptable variability and a similar average effect (Section 6.1.4).

7.3 Guidance to Regulatory Authorities

Guidance to Regulatory Authorities Related to Determining Reasonable Potential and Deriving Permit Limits:

- Regulatory authorities should use EPA's recommended statistical approach in deriving permit limitations. The statistical approach outlined in the TSD represents an effective and appropriately protective approach to effluent limit development (Section 6.1 and Appendix G.1).

- Regulatory authorities should calculate the facility-specific CV using point estimate techniques to determine the need for and derive a permit limit, even if the self-monitoring test results will be determined using hypothesis test procedures (Sections 3.4.1 and 6.2).

- Regulatory authorities that need to cite a characteristic CV for a promulgated method may use Tables A-1 and A-2 in Appendix A, which show the median CV from Tables 3-2 through 3-4, pending completion of between-laboratory studies.

- EPA recommends that regulatory authorities implement a step-wise approach to address toxicity. This approach can determine the magnitude and frequency of toxicity and appropriate follow-up actions for test results that indicate exceedance of a monitoring trigger or a permit limit (Section 6.5).

Guidance to Regulatory Authorities Related to Collecting Effluent Samples, Conducting the Toxicity Test, and Evaluating the Effect Concentrations:

- Regulatory authorities should design a sampling program that collects representative effluent samples to fully characterize effluent variability for a specific facility over time. At least 10 samples are needed to estimate a variance or CV with acceptable precision for a specific facility (Sections 6.1.3 and 6.2).

- Regulatory authorities should ensure that statistical procedures and test methods have been properly applied to produce WET test results. Evaluating other factors and data, such as biological and statistical quality assurance, and ensuring that test conditions and test acceptability criteria (TAC) have been met would be prudent (Sections 5.2 through 5.5).

- Regulatory authorities should apply both the upper and lower bounds using the PMSD as an additional TAC (Section 6.4 and Table 3-6). The State of North Carolina implemented an effective WET program that required additional TAC and guidance for test methods that served to minimize test method variability (Appendix F).

- Regulatory authorities should develop a QC checklist to assist in evaluating and interpreting toxicity test results (Section 5.3.1.1). See Appendix E for examples of State WET implementation procedures.

- Regulatory authorities should consider participation in the National Environment Laboratory Accreditation Program and should conduct routine performance audit inspections to evaluate individual laboratory performance. Inspections should evaluate the laboratory's performance with QC control charts based on reference toxicants, examine procedures for conducting the toxicity test procedures, and examine procedures for analyzing test results (Section 5.3.1.1).

- Regulatory authorities should incorporate revised technical guidance recently published by EPA captioned "Method Guidance and Recommendations for Whole Effluent Toxicity (WET) Testing"

(40 CFR Part 136) (USEPA 2000a). The guidance addresses: (1) error rate assumption adjustments; (2) concentration-response relationships; (3) incorporation of confidence intervals; (4) acceptable dilution waters for testing; (5) guidance on blocking by parentage for the chronic *C. dubia* test method; and (6) procedures for controlling pH drift.

7.4 Future Directions

- An independent peer-reviewed workshop should be convened to evaluate alternatives to the statistical approaches currently used in EPA's WET test methods. Such a workshop might suggest alternatives regarding (1) WET statistical flowcharts, (2) WET statistical methods used to estimate effect concentrations, and (3) test data interpretation and review guidelines (Section 5.5).

- Such a workshop might also evaluate additional QC requirements and recommendations regarding the specification of a reference toxicant and the concentrations to be tested for each test method (Section 5.3.1).

This page intentionally left blank.